A Semantic Reasoning Engine for Lifestyle Profiling in Support of Personalised Coaching

Sara Mikolajczak, Tom Ruette, Elena Tsiporkova
Data Innovation Group
ICT & Mechatronics, Sirris
Brussels, Belgium
email: {tom.ruette, elena.tsiporkova}@sirris.be

Milena Angelova, Veselka Boeva
Computer Systems & Technologies Department
Technical University of Sofia
Plovdiv, Bulgaria
email: milenaangelova89@gmail.com, vboeva@tu-plovdiv.bg

***Abstract*—The aim of this research is to develop a sound methodology enabling the profiling of coachees that are being supported by a lifestyle coach. This profiling is to be useful for the professional coaches to personalize their approach for helping people acquire and maintain a healthier lifestyle. For the coachee profiling, the known relevant factors for a healthy lifestyle are gathered, e.g., sleep duration and activity level. Based on coachee input in a questionnaire, to be filled in with support of the coach, a qualitative profile is generated in terms of food and liquids consumption, physical activity, personality and cardiovascular risk. The modeling and reasoning environment is realized with semantic technologies. We find that semantic technology is an efficient technology for performing user profiling in a digital coaching context.**

Keywords—coachee profile; ontology; Semantic Web; coachee modeling; Jena rules; protégé

I. INTRODUCTION

This research has been performed within a large ARTEMIS EU project WITH-ME [18], which aims to provide a digital coaching platform that continuously monitors, advises and interacts with coachees to help them acquire and maintain a healthier lifestyle. To reduce the variability of coachees, an initial coachee profile is constructed, based on answers to a questionnaire that is filled in by the coachee, together with a personal coach during the first coaching session. The part of the WITH-ME platform that generates an initial coachee profile is technically realized as an ontological model with a reasoning layer. This initial profile is used by the coach to determine the best strategy for helping the coachee acquire and maintain a healthier lifestyle.

Ontologies have been proven to be effective means for modeling user context [1]. Ontologies model concepts and relationships in a high level of abstraction, providing rich semantics for humans to work with and the required formalism for computers to perform mechanical processing and reasoning.

Using an ontology to model a user profile has already been proposed in various applications like web search [9], [12] and personal information management [7]. Up to this point, ontologies modeling user profiles are application-specific. Namely, each one has been created specifically for a particular domain. Taking into account the continuing incorporation of ontologies in new applications, there is an emerging need for a standard ontology to model user profiles. Such a standard ontology would facilitate the communication between applications and will serve as a reference point when profiling functionalities need to be developed.

In this work, however, we present an ontology for modeling coachee profiles in the domain of medical services including a variety of areas such as patient care, clinical and administrative decisions, assisting devices and patient diagnostics. The developed ontology will be put to use in a web-based software application that accepts input data from a form-based interface and then presents a coachee profile in the field of healthcare. The ontology development process starts with the requirements analysis phase where concepts, attributes, relationships and axioms are identified. In the design phase, a consistent conceptual model is defined over a set of tasks, which increases the complexity of the ontology step by step. In the development phase, a suitable ontological language is used to formalize the ontology, which can help to update the ontology according to the domain concepts in the maintenance phase.

In the remainder of this paper, we first discuss related work for applying ontologies in the medical domain (Section II). Then, we discuss a number of practical issues one can encounter during the creation of an ontology (Section III). Finally, we describe the software and results in some detail (Section IV). The paper ends with a conclusion (Section V).

II. RELATED WORK

Generally, an ontology is defined as a formal, explicit specification of a shared conceptualization [13]. An ontology describes the concepts and relationships that are important in a particular domain, providing a vocabulary for that domain as well as a computerized specification of the meaning of terms used in the vocabulary [13]. The aim of an ontology is to formalize domain knowledge in a generic way and to provide a common understanding of a domain, which may be used and shared by applications and groups. An ontology is also an emerging technology for knowledge representation [14], with applications for reasoning to infer new knowledge, and with applications for data integration [15].

An ontology consists of classes, properties and individuals. A class defines a concept; individuals realize a class; individuals are linked to each other via properties; individuals are linked to data values via properties. The use

 ISBN: 978-1-61208-424-4

of the term "individual" might seem confusing, since in natural language, it refers to a person. However, in the remainder of this paper, we will use "individual" consistently to refer to an instantiation of a class in an ontology.

Ontological technology has been widely adopted in the business and scientific communities as a way to share, reuse and process domain knowledge. Ontological technology is also central to many applications in fields including information management, systems integration and semantic web services [16]. Many studies have also demonstrated that ontologies are essential for the development of knowledge-oriented systems.

The most widely used terminology resources in the biomedical domain are the Gene Ontology and the Unified Medical Language System. The Gene Ontology (GO) project [19] is a collaborative effort to address two aspects of information integration in several domains of molecular and cellular biology. Namely, the GO project provides consistent descriptors for gene products in different databases and standardizes classifications for sequences and sequence features. The Unified Medical Language System (UMLS) [21] is maintained by the U.S. National Library of Medicine in order to facilitate the development of computer systems in the field of biomedicine and health. The UMLS consists of files and software that bring together many health and biomedical vocabularies and standards to enable interoperability between computer systems.

Huang et al. [4] developed a medical ontology to serve as the foundation for an intelligent Chinese Medical Diagnostic System (CMDS), which acts as a human expert to diagnose a number of digestive system conditions including stomachache, vomiting, hiccups, diaphragmatitis, diarrhea, dysentery, constipation, jaundice, tympanites, etc. In [5], Bouamrane et al. presented the design and implementation of an ontological knowledge-based pre-operative assessment support system, which is a generic clinical screening process, intended to identify early in a patient's journey the potential risk of complications during or after surgery. Juarez et al. proposed in [13] an ontology-based medical knowledge base called the Causal and Temporal Knowledge Acquisition (CATEKAT2), which provides physicians with a broad spectrum of medical knowledge. Kola et al. developed an ontological knowledge base to drive an Occupational Health Application (OCHWIZ), which provides suggestions as to possible causes and industries associated with a given clinical finding related to a specific occupation [15]. Based on these suggestions, it is also capable of inferring other diseases and conditions to watch out for. Alexandrou et al. described an ontological software platform SEMPATH, which can offer personalized treatment plans by using and managing health care business processes (clinical pathways) [14]. During the execution of clinical pathways, the system considers the patient's clinical status and reaction to the treatment scheme according to the Semantic Web Rule Language (SWRL) [29] rules in reconfiguring the next treatment steps.

Other important works with clinical relevance are SNOMED Clinical Terms (CT) [22] and ISO/EN 13606 Electronic Health Record (EHR) communication paradigm [23]. SNOMED CT [20] provides a consistent information interchange and is fundamental to an interoperable electronic health record. It allows a consistent way to index, store, retrieve, and aggregate clinical data across specialties and sites of care. It also helps in organizing the content of EHR systems by reducing the variability in the way data is captured, encoded and used for clinical care of patients and research. EHR Archetypes are formal and standardized specifications for the representation and organization of clinical information inside electronic health records. ISO/EN 13606-2 incorporates the "openEHR" archetype approach as a standard information model, and an exchange representation, for the communication of electronic health record archetypes [23].

III. Ontology Creation Issues

A. Application Context

The goal of this research is to develop an application capable of generating a complete and qualitative lifestyle profile based on the input of a coachee. An initial ontology model together with the reasoning layer implemented in SWRL has been conceived and validated as a part of the Master Thesis work of Sara Mikolajczak at the KU Leuven, Belgium.

Lifestyle can be modelled taking into account different aspects, e.g., eating and drinking habits like calorie and nutrient intake, activity levels, body and weight levels like BMI (Body Mass Index), WC (Waist Circumference) and WHR (Waist-Hip Ratio), levels of motivation and levels of general (un)healthy habits (risk factors) like smoking and sleeping habits.

To collect the relevant factors that influence a healthy lifestyle, an extensive study of the literature [16][17] has been performed as part of a master thesis. From this literature study, important life style areas have been extracted with relevant data, such as boundary values and possible conditions that may be the consequence of exceeding these boundaries. These data have been used in the construction of the coachee profile.

For some areas, it was not possible to find scientifically validated specifications of the boundary values. In this case, non-scientific values were used to provide a full proof-of-concept implementation. Therefore, further research is required to refine these values in order to provide a correct and justified tool to the coach and coachee.

B. Definition

An ontology is a "hierarchal structuring of knowledge about concepts by sub-classing them according to their properties and qualities" [6]. It can also be defined as "a declarative model of a domain that defines and represents the concepts existing in that domain, their attributes and the relationships between them" [2][6]. Thus, our lifestyle ontology gives a formalization of concepts (e.g., person, habit, motivation) in their respective classes (e.g., Person, Habit, Motivation) and the relations (e.g., is a Person, has a

Habit, has a Motivation) that can exist between them. A correct definition of classes is very important for data sharing and knowledge representation.

C. Classification

Ontologies can be classified according to the level of detailed knowledge they provide. Upper ontologies provide very generic knowledge with low domain specific knowledge. For example, a disease ontology is an upper ontology compatible for any biomedical domain. General ontologies represent knowledge at an intermediate level of detail independently of a specific task. Domain ontologies represent knowledge about a particular part of the world, such as medicine, and should reflect the underlying reality through a theory of the represented domain. Finally, ontologies designed for specific tasks are called application ontologies.

D. Description

We have defined an ontology as a specification and formalization of concepts and relations between them. The domain concepts are represented by “classes”. The features of a concept are described by “properties”. These properties can be either relationships between classes, or data values of “individuals”, which are the instantiations of a class. Together with these “individuals”, this is what constitutes the domain knowledge base.

Classes are the main focus in an ontology. Classes can be sub-classed to describe more specific features of a class. For example, if we define a class Habit, it includes all the habit classes in the habit domain. The Habit class can be sub-classed to specify more specific habits like ConsumptionHabit or SleepingHabit.

Properties can be created to describe properties of a class or individual. For example, we can define a property of a class named “hasSleepingHabit”, which connects a Habit Class to the SleepingHabit class. An individual that instantiates the SleepingHabit class can have a property sleepingHabit, which holds a data value that gives the amount of hours somebody typically sleeps. We can also mention that it is possible to store a data value range as a property of an individual, e.g., sleepingHabit could contain a range of the amount of hours somebody sleeps. Moreover, fuzzy logic reasoners can be applied to optimally leverage the concept of a range.

Figure 1 shows a summary of the above discussion. The description of an ontology domain includes [3]:

- The definition of a concept in the domain as classes.
- The definition of an instantiation of the class as individuals.
- The definition of attributes of classes or individuals as properties.

E. Ontology structure

The ontology model contains three ontologies in layers: a Core Ontology Layer, a Profiling Ontology Layer and a Reasoning Ontology Layer. They are depicted in Figure 2.

1. *Core ontology* – is the main ontology, which contains the classes and properties. It also contains a Standard_Profile, which is an individual that contains the boundary values that were recorded during the literature review phase.
2. *Profiling ontology* – contains multiple individuals that describe different parts of the final coachee profile. The final coachee profile is divided into an “eating and drinking profile”, a “motivation profile”, a “physical profile” and a “health profile”. These four aspects are discussed in some more detail below.
3. *Reasoning ontology* – This layer contains the individuals that are to be populated with coachee data.

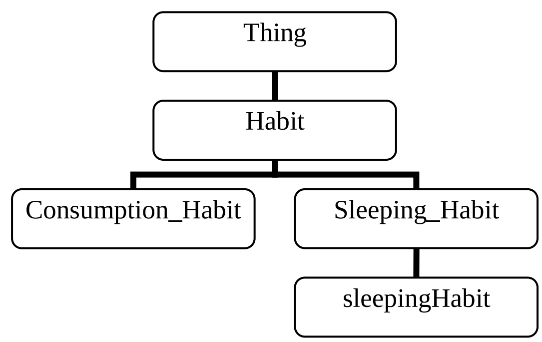

Figure 1: The amount of hours that a coachee habitually sleeps is stored in the ontology.

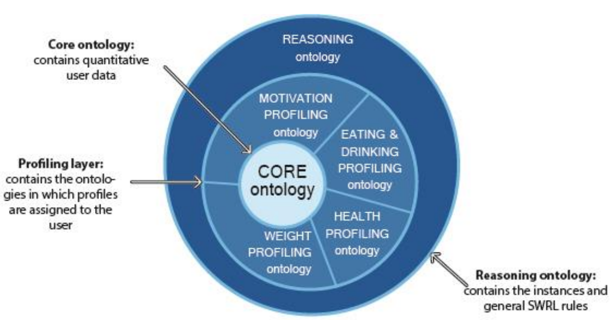

Figure 2: Ontology structure.

Below, we discuss in some more detail the four aspects of the Profiling ontology layer:

Eating and Drinking Profile - In this part of the ontology, the daily consumed food is evaluated. The evaluation of calorie intake is based on the daily calorie requirement (DCR), which in turn is based on the activity level in combination with the basal metabolic rate (BMR). The evaluation of protein, fat and carbohydrate intake is based on the recommended rate of each nutrient in the total amount of calories.

 ISBN: 978-1-61208-424-4

Motivation profile - Based on answers given on a list of multiple choice questions, the coachee receives a score for being intrinsically and extrinsically motivated. The reasoner evaluates these scores and returns an optimal approach for motivating the coachee.

Physical profile - The physical profile is further split into two parts: the first part concerns the body profiling, the second part the activity profiling. The body profile consists of a body mass index level, a waist-hip ratio level and a waist circumference level. These levels indicate a risk for a cardiovascular disease, given a certain BMI, WHR and WC. The activity profile contains five activity levels going from "sedentary" to "extremely active". The assignment of an activity level to a coachee is based on the her job(s), hobbies and means of transportation.

Health profile - In this part of the ontology, the sleeping and smoking habits are modeled. The optimal daily number of hours of sleep is defined according to certain age categories, e.g., the optimal number of daily hours of sleep for younger people is higher than for older people. The smoking habit class is defined by the daily amount of tobacco products being consumed.

F. Rules

Two formalization approaches for expressing rules have been investigated in this paper. First, the Semantic Web Rule Language (SWRL) can be used to create rules in an ontology. The SWRL is a proposed language for the Semantic Web that can be used to express rules as well as logic by combining Web Ontology Language (OWL) Description Logic (DL) or OWL Lite with a subset of the Rule Markup Language. Second, rules can also be implemented in the format specific for Apache Jena [28], which is a Java-oriented system for handling ontologies. Jena provides an Application Programming Interface (API) to extract data from and to write to Resource Description Framework (RDF) [24] graphs. Since the syntax of Jena rules is more readable, and since these rules can be more easily integrated in a software production environment, we preferred the Jena approach over the SWRL approach.

One can see below a rule that is written in Jena syntax. This rule calculates the DCR of a coachee. DCR is the amount of calories the coachee should eat in a day to maintain her current weight at the current height and age. It is calculated based on the coachee’s level of Basic Metabolic Rate (BMR) and how active the coachee is:

```
[DCR:
        (?p rdf:type core:Person)
        (?p physical:hasActivityProfile ?activityProfile)
        (?activityProfile physical:activityValue ?activityValue)
        (?p core:hasExternalExaminationResult ?BMRCalc)
        (?BMRCalc core:examinationType 'BMR')
        (?BMRCalc core:externalValue ?BMR)
        product(?activityValue, ?BMR, ?DCR)
        (?p core:hasExternalExaminationResult ?DCRCalc)
        (?DCRCalc core:examinationType 'DCR')
        -> (?DCRCalc core:externalValue ?DCR)
]
```

IV. DESCRIPTION OF SOFTWARE AND RESULTS

The software implementation that we used to build a proof-of-concept for the WITH-ME project [18] adopts a decoupled architecture, where the evaluation of the data values by the ontology and its rules is strictly separated from the web-based interface. These two components communicate via a so-called messaging system, where the output of the first component (the interface) becomes the input for the second component (the ontology and rules).

The web interface contains a number of questions that need to be filled in, some of which are depicted in Figure 3.

The decoupled architecture of the web-based software requires that the data, i.e. answers to the questionnaire, from the web-based interface is communicated to the separated ontology processing backend. To communicate between these two components, a shared message format is used, i.e. Java Script Object Notation (JSON) [25]. JSON is a lightweight data-interchange format, that is easily readable for humans, and easily generated and parsed by machines. The messaging system that is being used in WITH ME is Microsoft Azure Service Bus [26].

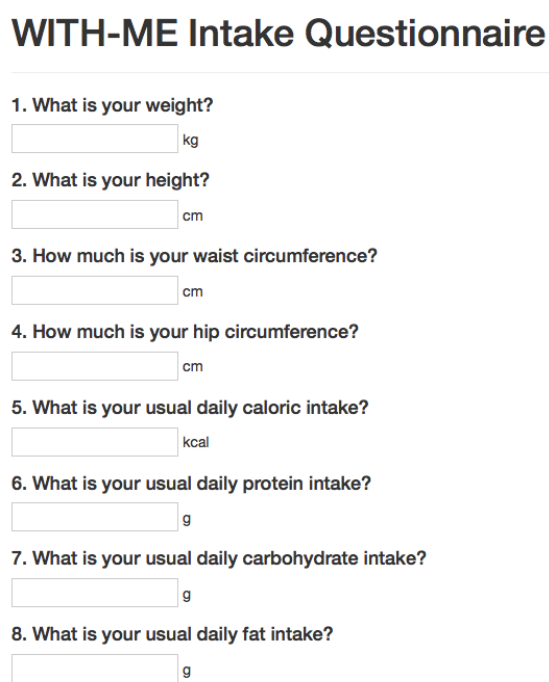

Figure 3: A screenshot of the web-based questionnaire interface.

The collected personal data are sent to the backend, which consists of a Apache Jena implementation around the ontology. The backend accepts the input data and stores them as properties in the appropriate individuals of the ontology model. A Pellet reasoner [27] is then used in the developed system. Pellet is an open-source Java-based OWL 2 reasoner. The Pellet reasoner analyzes the received data from the web system and generates newly inferred data. The inferred data together make up the coachee profile. The output of the backend is again a JSON message, which is

 ISBN: 978-1-61208-424-4

being submitted to the messaging system. From there, it can be picked up by another decoupled component that could, as an example, generate visualizations of the coachee profile.

The proposed profiler will be evaluated on multiple coachees during the pilot phase of the WITH ME project, starting September 2015.

V. CONCLUSION AND FUTURE WORK

This paper has presented an approach for designing a semantic reasoning engine for coachee profiling. An ontological approach to knowledge representation in healthcare has been selected in the coachee profile design.

The developed system uses a web-based interface for collecting coachee data and an ontology to analyze and process the entered coachee data. The performed analysis generates a coachee profile as a result. This profile can be used to optimize the coaching activities of a professional lifestyle coach, or as a basis for the creation of other software applications that expand the applicability of the developed one. For the former, an example is that a professional lifestyle coach can choose a motivational style on the basis of the motivation profile that is generated by the profiler. For the latter, an example is that a software application can use the profile to automatically recommend objectives to the coachee, on the basis of objectives of other coachees with similar user profiles.

REFERENCES

[1] N. Gurano, "Formal Ontology and Information Systems", Formal Ontology in Information Systems, Proceedings of FOIS'98, Trento, Italy, 6-8 June 1998, pp. 3-15.

[2] J. C. Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[3] N. F. Noy and D. L. McGuinness. "Ontology Development 101: A Guide to Creating Your First Ontology". Stanford Knowledge Systems Laboratory Technical Report KSL-01-05M.

[4] J. Huang and M. Y. Chen, "Integrated design of the intelligent web-based Chinese Medical Diagnostic System (CMDS) – Systematic development for digestive health", Expert Systems with Applications, vol. 32, January 2007, pp. 658-673.

[5] M. M. Bouamrane, A. Rector, and M. Hurrell, "Using Owl ontologies for adaptive patient information modelling and preoperative clinical decision support", Knowledge and Information Systems, vol. 29, no. 2, November 2007, pp. 405-418.

[6] T. R. Gruger, "Toward Principles for the Design of Ontologies Used for Knowledge Sharing", International Journal of Human-Computer Studied, vol. 43, no. 5-6, November-December, 1995, pp. 907-928.

[7] V. Katifori, A. Poggi,. M. Scannapieco, T. Catarci, and Y. Ioannidis, "OntoPIM: how to rely on a personal ontology for Personal Information Management", In Proc. of the 1st Workshop on The Semantic Desktop, 20015, pp. 78-81.

[8] M. K. Smith, C. Welty, and D. L. McGuinness. "OWL Web Ontology Language Guide. W3C Recommendation", 2004, http://www.w3.org/TR/owl-guide/
[retrieved: 6, 2015]

[9] J. Trajkova and S. Gauch, "Improving Ontology-based User Profiles", Proc. of RIAO 2004, University of Avignon (Vaucluse), France, April 26-28, 2004, pp. 380-389.

[10] Y. Tang and R. Meersman, "DIY-CDR: an ontology-based, Do-It-Yourself component discoverer and recommender," Personal and Ubiquitous Computing, vol. 16, no. 5, June 2011, pp. 581-595.

[11] A. Smirnov, N. Shilov, T. Levashova, L. Sheremetov, and M. Contreras, "Ontology-driven intelligent service for configuration support in networked organizations," Knowledge and Information Systems, vol. 12 no. 2, July 2007, pp. 229-253.

[12] E. Rich, "Users are individuals: individualizing user models", International Journal of Man-mashine Studies 18(3), 1983, pp. 199-214.

[13] J. M. Juarez. T. Riestra, M. Campos, A. Morales, J. Palma, and R. Marin, "Medical knowledge management for specific hospital departments", Expert System Application, vol. 36, no. 10, December 2009, pp. 12214-12224.

[14] D. A. Alexandrou, I. E. Skitsas, and G. N. Mentzas, "A Holistic environment for the Design and Execution of Self-Adaptive Clinical Pathways", IEEE Trans. On Information Technology in Biomedicine, vol. 15, no. 1, January 2011, pp. 108-118.

[15] J. Kola, B. Wheeldin, and A. Rector, "Lessons in building OWL Ontology driven applications: OCHWIZ – an Occupational Health Application", in Proc. e-Scinece All Hands Conf., UK, 2007, pp. 71-78.

[16] NIH (National Heart, Lung and Blood Institute, "What Are the Health Risks of Overweight and Obesity?", 2012, http://www.nhlbi.nih.gov/health/health-topics/topics/obe/risks.html
[retrieved: 6, 2015]

[17] P. Kopelman, „Health risks associated with overweight and obesity," obesity reviews, vol. 8, nr. 1, 2007, pp. 13-17.

[18] http://www.with-me-project.eu/
[retrieved: 6, 2015]

[19] Gene Ontology Consortium, "The Gene Ontology (GO) Database and Informatics Resources", Nucleic Acids Research, Database issue, vol. 32, 2004, pp. D258-D261

[20] P. Ruch, J. Gobeill, C. Lovis, and A. Geissbühler, "Automatic medical encoding with SNOMED categories", BMC Medical Informatics and Decision Making, vol.8, 2008, supplement 6.

[21] http://www.nlm.nih.gov/research/umls/
[retrieved: 2015-06-01]

[22] http://www.ihtsdo.org/snomed-ct/
[retrieved: 6, 2015]

[23] http://www.eurorec.org/services/archetypes/
[retrieved: 6, 2015]

[24] http://www.w3.org/TR/2014/REC-rdf-syntax-grammar-20140225/
[retrieved: 6, 2015]

[25] http://json.org/
[retrieved: 6, 2015]

[26] http://azure.microsoft.com/en-us/services/service-bus/
[retrieved: 6, 2015]

[27] E. Siring, B. Parsia, B. C. Grau, A. Kalyanpur, Y. Katz, „Pellet: a practical OWL-DL reasoner", Web Semantics: Science, Services and Agent on the Word Wide Web, vol.5, June 2007, 51-53.

[28] http://jena.apacha.org
[retrieved: 6, 2015]

[29] http://www.w3.org/Submission/SWRL/
[retrieved: 6, 2015]